NBER Digitization Tutorial Application
Daniel Yue
December 28th, 2022

Contact Information

Name: Daniel Yue
PhD Program: Technology & Operations Management (Harvard Business School)
Email: dyue@hbs.edu
Cell-phone: 202-704-6446
Office: Morris Hall 101A

Description of Graduate Studies

Graduate Program:

Harvard Business School, 2019 to present (*expected graduation in 2024*)
Ph.D. Candidate in Business Administration
Field: Technology & Operations Management (Innovation Track)

Coursework:

Economics

1. Microeconomic Theory I (Maciej Kotowski)
2. Microeconomic Theory II (Chris Avery and Elon Kohlberg)
3. [MIT] Organizational Economics (Bob Gibbons and Charles Angelucci)
4. NBER Entrepreneurship Research Bootcamp (David Robinson)

Methods

1. Principles of Econometrics (Elie Tamer)
2. Econometric Methods (Isaiah Andrews)
3. Design of Field Research Methods (Robin Ely)
4. Field Experiments (Shawn Cole)

Technology Management

1. Empirical Technology & Operations Management (Feng Zhu)
2. [MIT] The Economics of Information and IT (Erik Brynjolfsson)
3. The Economics of Innovation and Digitization (Shane Greenstein)

Strategic Management

1. Organizational Theory (Frank Dobbin)
2. Foundations of Strategy (Juan Alcacer)
3. Advanced Topics in Strategy (Juan Alcacer)

Teaching:

HBS 2231 — From Data To Decisions: Leveraging Experiments for Effective Strategy, Marketing, and Entrepreneurship (Teaching Assistant for Prof. Michael Luca)

Brief Summary of Research Interests

My research interests focus on questions about how firms leverage data analytics to create competitive advantage. In particular, my dissertation will focus on how businesses develop and use artificial intelligence to improve business operations and create value. I am in the process of forming my dissertation committee; my committee chair will be Prof. Shane Greenstein. I hope it will comprise Prof. Karim Lakhani and Prof. Kyle Myers.

Artificial intelligence is a statistical framework where firms leverage large amounts of data to generate predictions that are useful for business operations. For example, what is the best product to recommend to my customer? What department is best suited to handle my customer's service request? What historical company documents are most relevant to my current project? Business interest in artificial intelligence exploded following from the 2012-on breakthroughs in the field of deep learning, but, even a decade later, there are many questions to explore regarding how to best develop and use of the technology. To this point, I have pursued research projects focusing on: A. how the prominent involvement of companies impacts AI research, B. the complementary tools and skills needed for firms to effectively develop AI applications, and C. the impact of open-sourcing and openness norms on the pace of research and adoption of AI techniques. I have also co-authored a teaching case about a prominent open-source AI startup (Hugging Face).

In particular, in my (likely) job market paper, I develop empirical strategies to study effect that a company hiring a researcher has on that researcher's rate and direction of research. Leveraging a large-scale bibliometric data set and controlling for researcher selection into corporate involvement, I find that companies affect the direction of research but not the rate of publication for hired researchers. I relate these findings to policy questions around the creation of national resources for AI resources, as well as theoretical and managerial questions about the role of corporations in science and the business strategies that benefit from such involvement.

Beyond AI, my interests extend to the development and use of data analytics in businesses more broadly. This includes prior (unpublished) projects about experimentation in firms enabled by digitization (e.g. A/B testing). More generally, I am excited by projects that study the characteristics of the ecosystem from which they transformative technologies emerge, and the factors that enable firms to leverage these technologies, learn from their environments, and take better actions in light of what is true about the world around them.

Research Projects

Active Projects

I, Google: Estimating the Impact of Corporate Involvement on AI Research (solo-authored) [draft in progress]. Working Abstract: *While modern corporate funding of scientific research is an indisputable fact, the impact of corporate involvement on scientific progress is controversial. Corporate influence may lead to companies redirecting research activities or absorbing efforts into non-research activities in a way that benefits the company but harms overall scientific progress. However, corporations also provide resources, data sets, collaborators, and real-world problems they may otherwise have not been able to access or know about, spurring knowledge creation. This paper empirically assesses the impact of corporate involvement on AI research by exploiting aspects of the "brain-drain" of AI researchers to companies during the past decade. Controlling for the effect of researcher selection into corporate involvement, we find that corporate involvement does not reduce the rate or impact of researcher's published discoveries. However, corporate involvement does significantly alter the direction of their research away from disciplinary work and towards work that is closer to the commercial interests of the company.*

Nailing Prediction: Experimental Evidence on The Value of Tools in Predictive Model Development (with Iavor Bojinov and Paul Hamilton) [working paper available]. Abstract: *Predictive model development is understudied despite its importance to modern businesses. Although prior discussions highlight advances in methods (along the dimensions of data, computing power, and algorithms) as the primary driver of model quality, the value of tools that implement those methods has been neglected. In a field experiment leveraging a predictive data science contest, we study the importance of tools by restricting access to software libraries for machine learning models. By only allowing access to these libraries in our control group, we find that teams with unrestricted access perform 30% better in log-loss error — a statistically and economically significant amount, equivalent to a 10-fold increase in the training data set size. We further find that teams with high general data-science skills are less affected by the intervention, while teams with high tool-specific skills significantly benefit from access to modeling libraries. Our findings are consistent with a mechanism we call `Tools-as-Skill,' where tooling automates and abstracts some general data science skills but, in doing so, creates the need for new tool-specific skills.*

Conference Papers and Teaching Cases

How Open Source Machine Learning Software Shapes AI (with Maximilian Langenkamp) [published], *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*.

Hugging Face: Serving AI on a Platform (with Shane Greenstein, Kerry Herman, and Sarah Gulick) [published], *HBS Teaching Case, 2022*.

- Taught in Professor Shane Greenstein's Fall 2022 MBA Class HBS 2134 "Digital Innovation and Transformation".